



Lloyd, C., Freer, J., Johnes, P., Coxon, G., & Collins, A. (2015). Discharge and nutrient uncertainty: Implications for nutrient flux estimation in small streams. *Hydrological Processes*, 30(1), 135–152. <https://doi.org/10.1002/hyp.10574>

Peer reviewed version

Link to published version (if available):
[10.1002/hyp.10574](https://doi.org/10.1002/hyp.10574)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

1 **Discharge and nutrient uncertainty: implications for nutrient flux estimation in small streams.**

2 **Lloyd, C. E.M.^{1*}, Freer, J.E.¹, Johnes, P. J.¹, Coxon, G.¹ and Collins, A.L.².**

3 ¹ School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK.

4 ² Sustainable Soils and Grassland Systems Department, Rothamsted Research – North Wyke,
5 Okehampton, EX20 2SB, UK.

6
7 * Corresponding author: Charlotte Lloyd

8 Address: School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS.

9 e-mail: charlotte.lloyd@bristol.ac.uk

10 Telephone: +44 (0)117 9288808 Fax: +44 (0)117 928 7878

11 Article accepted for publication in Hydrological processes and available online

12 DOI: 10.1002/hyp.10574

Abstract

The measurement of discharge is fundamental in nutrient load estimation. Due to our ability to monitor discharge routinely, it is generally assumed that the associated uncertainty is low. This paper challenges this preconception, arguing that discharge uncertainty should be explicitly taken into account to produce robust statistical analyses. In many studies, paired discharge and chemical datasets are used to calculate 'true' loads and used as the benchmark to compare with other load estimates.

This paper uses two years of high frequency (daily and sub-hourly) discharge and nutrient concentration data (nitrate-N and total phosphorus (TP)) collected at four field sites as part of the Hampshire Avon Demonstration Test Catchment (DTC) programme. A framework for estimating observational nutrient load uncertainty was used which combined a flexible non-parametric approach to characterising discharge uncertainty, with error modelling that allowed the incorporation of errors which were heteroscedastic and temporally correlated.

The results showed that the stage-discharge relationships were non-stationary, and observational uncertainties from ± 2 -25% were recorded when the velocity-area method was used. The variability in nutrient load estimates ranged from 1.1-9.9% for nitrate-N and from 3.3-10% for TP when daily laboratory data were used, rising to a maximum of 9% for nitrate-N and 83% for TP when the sensor data were used. However, the sensor data provided a better representation of the 'true' load as storm events are better represented temporally, posing the question: is it more beneficial to have high frequency, lower precision data or lower frequency but higher precision data streams to estimate nutrient flux responses in headwater catchments?

Keywords: Discharge, nutrients, uncertainty, load estimation, non-parametric, stage-discharge, velocity-area

1. Introduction

Stream discharge monitoring is fundamental to our characterisation of catchment responses. Historical records for discharge are normally longer than those available for other variables, and at a higher temporal resolution, providing the longest and most detailed record of changes in catchment function over time. Discharge plays an integral role in analyses such as nutrient load estimation and in understanding the processes driving nutrient delivery to streams and transport downstream as those loads are delivered from land to sea. Due to our ability to monitor stream discharge routinely, usually at high temporal resolution (typically 15 or 30 minute), there seems to be a general view that the associated uncertainties are low or non-existent (Moatar *et al.*, 2013; Tomkins, 2014). However, discharge is rarely measured directly and instead, a time series of river water levels (and sometimes velocity) are collected at a gauging station and converted to a corresponding time series of discharge. This relationship between stage, velocity and discharge is subject to a number of different sources of observational uncertainties which can influence the quality of discharge data being collected (Hamilton and Moore, 2012; McMillan *et al.*, 2012) including errors in observations of stage or velocity, instability of channel cross-sections, seasonal vegetation growth and clearance, and uncertainties related to the generation of a stage-discharge rating curve, particularly where out-of-bank flow is a regular occurrence. Therefore, we argue that it is important that discharge uncertainty is taken into account explicitly in any study exploring catchment responses, whether induced by natural environmental variability, climate change or direct human manipulation of the water cycle.

Recently, there has been an increase in the number of integrated research platforms set up internationally to monitor hydrochemical fluxes from headwater or small tributary sub-catchments in order to assess the efficacy of on-farm mitigation strategies to reduce water pollution from point and diffuse sources. A number of these studies have made use of novel sensor and bankside analyser technologies to deliver high-frequency measures of the concentrations of nutrient species and other variables such as instream temperature and dissolved oxygen, in conjunction with discharge measurements. An example is given in Outram *et al.* (2014) who examined storm flow responses in stream nutrient chemistry at three UK sites using high frequency bankside analysers and sensor technologies generating 15 minute resolution discharge measurements and 30 minute resolution measurements of nitrate and total phosphorus concentrations. In each example, the data are used to investigate how nutrient loads are varying with time or how load dynamics are affected by changes in storm duration and intensity, often with the overall aim of informing appropriate land management strategies for reducing nutrient export from land to stream. Uncertainty analyses associated with this type of work have usually focused on errors based on the method (e.g. interpolation or extrapolation)

of load estimation (Rekolainen *et al.*, 1991; Kronvang and Bruhn, 1996; Webb *et al.*, 2000; Johnes, 2007; Horowitz, 2008; Krueger *et al.*, 2009) or on the impact of sampling frequency on load estimation (Johnes, 2007; Horowitz, 2008; Jordan and Cassidy, 2011; Halliday *et al.*, 2012; Krueger *et al.*, 2012). For example Johnes (2007) investigated the impact of load estimation methodology, sampling frequency and catchment characteristics on the uncertainty in phosphorus loads in 39 UK rivers. In this and many similar studies (e.g. Horowitz, 2003; Moatar and Meybeck, 2007; Verma *et al.*, 2012) paired daily or sub-daily discharge and chemical datasets are used to calculate loads and are considered to be the 'truth', providing a baseline against which to compare coarser temporal resolutions of data using the same or contrasting load estimation methodologies. The work presented here provides a new approach which highlights the observational uncertainties associated with both discharge and chemical variables, which should be taken into explicit consideration during analyses, even when high temporal frequency paired measurements are available. This paper focuses on the impact of discharge uncertainties on the calculation of nutrient loads in headwater streams, and the combined impact of these with the observational nutrient uncertainties in the overall assessment of error.

Discharge in small streams is often derived from quasi-continuous stage measurements combined with a rating curve developed for the individual field site (ISO 1100-2, 2010). However, this method relies on there being a temporally stable relationship between stage and discharge, which is often not the case as this paper will show. Many recent papers have discussed rating curve uncertainty and attempted to quantify the effects on estimated discharge (e.g. Di Baldassarre and Montanari, 2009; Liu *et al.*, 2009; McMillan *et al.*, 2010; Jalbert *et al.*, 2011; Westerberg *et al.*, 2011; Birgand *et al.*, 2013; Juston *et al.*, 2014; Shao *et al.*, 2014). One of the most significant sources of uncertainty associated with the use of a rating curve results from extrapolation of the rating curve beyond the lowest or highest stage-discharge measurements (Rantz and al., 1982; Di Baldassarre and Montanari, 2009), which is often unavoidable if the periods of hydrological interest are extreme low or high flows, particularly for the latter where flows become out-of-bank. These periods can be especially important for nutrient related water quality studies where sediment-associated nutrient fractions are transported to and within the stream in the highest flow events. To avoid the uncertainties associated with the use of a rating curve, river discharge can be calculated using direct measurements of velocity and cross-sectional area (usually based on a known geometry and stage height measurements) known as the velocity-area method (ISO 1088, 2007). While continuous monitoring of multiple variables can be more costly and labour intensive, uncertainties are thought to be generally lower than using a rating curve methodology, especially in small headwater catchments where the systems are often temporally dynamic. That said, there are still important and potentially significant uncertainties

associated with the velocity-area method that should be acknowledged in scientific studies. A review of the sources of uncertainty associated with velocity-area calculation was provided by Pelletier (1988) based on 140 previous studies, identifying the main sources of uncertainty as the precision of the velocity measurements, the variability of the velocity across the river cross-section and the determination of the river channel geometry. More recently, a test of the applicability of an acoustic Doppler device for the measurement of water velocity was carried out by Nord *et al.* (2014) who concluded that the technique could be reliable under a variety of environmental conditions providing that the instruments were appropriately calibrated in the field. An in-depth review of river discharge uncertainty was presented by McMillan *et al.* (2012), who compiled evidence from numerous studies to illustrate the typical ranges of uncertainty from instrument accuracy and precision, rating curve uncertainty and newer more direct measurement techniques such as acoustic Doppler velocimetry (ADV). While there is now acknowledgement of discharge measurement uncertainty within both field hydrology and hydrological modelling research, few studies have incorporated estimates of discharge uncertainty within an assessment of nutrient loads. Those which have started to address this knowledge gap tend to be assessing the uncertainties associated with discharge rating curves in combination with nutrient load models (Kulasova *et al.*, 2012) rather than measurement uncertainty associated with velocity-area calculations. This paper addresses this gap in the international literature.

More specifically, this paper aims to assess the impact of discharge and nutrient observational uncertainties on nutrient load estimation by utilising two years of high frequency (daily and sub-hourly) discharge and nutrient concentration data (nitrate-N ($\text{NO}_3\text{-N}$) and total phosphorus (TP)) collected as part of the Hampshire Avon Demonstration Test Catchment (DTC) programme (McGonigle *et al.*, 2014; Outram *et al.*, 2014). These data provide quasi-continuous stage and velocity data at four field sites that can be used to examine the stability of the stage-discharge relationship over time. Therefore, the principle of stationarity in this relationship can be tested and/or any dominant deviations from this due to structural changes (for example, seasonality) taken into account. Contrary to many previous studies, the error in the discharge data is not assumed to be from a parametric distribution and temporally independent. Here, an error model is proposed to allow for heteroscedasticity and temporal autocorrelation in the observational datasets. In addition, the uncertainties associated with the measurement of discharge and the nutrient parameters were estimated and the impact of these uncertainties on the routine calculations of nutrient loads was quantified. By applying this nutrient load estimation framework across different catchment geologies, the sensitivity and importance of catchment hydrogeological function as an influence on estimation uncertainty has also been assessed.

2. Methodology

2.1 Site descriptions

The Hampshire Avon is classified as a chalk system since its upper reaches drain chalk, where headwaters are represented by winterbournes. However, the western headwaters drain clay, and sands and gravels dominate the lower reaches of the catchment. Land use in the catchment is predominantly rural (98%), comprising arable land, improved and rough pasture and semi-natural woodland. The location of the four sampling sites discussed in this study is shown in Figure 1, and the characteristics of the corresponding sub-catchments are summarised in Table 1. The hydrological regime of each of the sub-catchments is influenced by the geology and in this study two sub-catchments are considered which are predominantly underlain by clay (Priors Farm and Cool's Cottage in the Sem catchment) and two which are underlain by chalk (Brixton Deverill in the Wylve catchment and Ebbesbourne Wake in the Ebble catchment). Detailed descriptions and conceptual models of the hydrogeology of each of the study sub-catchments can be found in Allen *et al.* (2014), while a detailed description of the nutrient hydrochemistry behaviour of the Wylve catchment over a 2 year period can be found in Yates and Johnes (2013). The sub-catchments for this study were chosen to represent both of the main types of geology found in the Hampshire Avon river basin in order to investigate whether discharge uncertainty is greater in some lithologies compared to others with differing hydrogeological character.

2.2 Data collection

Discharge and nutrient concentration data ($\text{NO}_3\text{-N}$ and TP) were collected at each of the four sampling sites between March 2012 and March 2014.

2.2.1 Discharge data

Discharge data for the River Wyle at Brixton Deverill were obtained from an existing gauging station maintained by the Environment Agency (EA) (Gauge number 43806). A stage-discharge rating curve had been developed using manual gauging from the field site which was used in conjunction with 15 min resolution stage height data collected using a Thistle 24R Incremental Shaft Encoder with a float and counterweight, to calculate discharge. Since 2010, this method has been used by the EA only during modular flow conditions and when non-modular flows occur (mainly due to large quantities of weed growth during the growing season) a second gauge recording stage and velocity is used to calculate discharge using the velocity-area method, which allows a comparison in this paper of these different approaches to quantifying resultant loads. The concrete channel cross-section meant that only one initial measurement of the cross-sectional area was necessary as the section geometry was

fixed. The resultant discharge time series is derived from a combination of the two calculated data sets.

Discharge data for the River Sem at Priors Farm and Cool's Cottage were collected using a Mace Flow Pro for the measurement of velocity and stage at 15 min resolution, and discharge was calculated using the velocity-area method. For the River Ebbles at Ebbesbourne Wake, stage height and velocity measurements were collected at 15 min resolution using a NIVUS OCM F ultrasonic sensor. This system was used at the Ebbesbourne site since it is documented as performing better in the low flow conditions (NIVUS, 2013) characterising this winterbourne stream. The channel cross-sections were surveyed once in the course of the study. All of the cross-sections were considered stable due to the installation of new concrete structures, or the measurements being taken within an existing concrete-sided bridge or culvert (see Figure 2). Only at Priors Farm did the stage repeatedly exceed the height of the concrete section. When this occurred, the stream would overtop this structure. To account for this extra flow, a weir equation was implemented to estimate the additional volume of water flowing outside of the culvert during out-of-bank flows.

2.2.2 Nutrient data

Samples were collected daily using an ISCO 3700 autosampler at all field sites and were returned once a week to the laboratory to be analysed. A filtered sub-sample was analysed for total oxidised nitrogen (TON) using the hydrazine-copper reduction method and measured colourimetrically at 540 nm. Nitrite concentrations were negligible and therefore TON is referred to as $\text{NO}_3\text{-N}$ hereafter. An unfiltered sub-sample was analysed for total phosphorus via a persulphate oxidation process (Johnes and Heathwaite, 1992) followed by a reaction of the digestate with heptamolybdate and potassium antimony (III) oxide tartrate in acidic solution then a reduction with ascorbic acid. The solution was then analysed colourimetrically at 880 nm. In addition to the daily sampling campaign, on the River Wylfe at Brixton Deverill $\text{NO}_3\text{-N}$ and TP were measured bank-side at 30 min resolution using a Hach Lange Nitratex Plus SC and a Phosphax Sigma system, respectively. The Nitratex system measures nitrate in-situ as $\text{NO}_3\text{-N}$ using an UV optical sensor. The sensor was calibrated every 3 months according to the manufacturer's guidelines using standard solutions. The Phosphax Sigma is a wet chemistry analyser which uses reagents to perform a digestion at high temperature and high pressure, and then all the P compounds are measured as orthophosphate colourimetrically using the molybdate method. The Phosphax was automatically calibrated daily and reagents were changed every 3 months when the instrument was serviced.

2.3 Data analysis

The procedure for quantifying the measurement uncertainties within the discharge and nutrient datasets is illustrated in Figure 3 and described in detail in the sections below.

2.3.1 Quantifying uncertainties

Discharge observational uncertainty was estimated using a period of the data record where the stage-discharge relationship was stable, so that any resultant variability reflected measurement uncertainty (aleatory) rather than structural changes (epistemic uncertainties) to the system behaviour (see section 3.1). A stable period usually occurred during any given season. These periods represent times when it was believed that epistemic changes were not occurring due to other environmental factors, for example vegetation growth changing the shape of the stage-discharge relationship. Several different stable periods were analysed and the scale of scatter was similar. Therefore the longest stable period (always weeks to months) for each site was chosen for the analysis so that a representative estimation of the autocorrelation and heteroscedasticity could be gained from using the widest range of stage measurements. To estimate the discharge uncertainty, a non-parametric LOcal Weighted regrESSion (LOWESS) approach was used for constructing a best-fit rating curve and to calculate the resultant uncertainty bounds. The LOWESS procedure takes each stage-discharge measurement as a central value and selects a window incorporating a chosen number of data points (a span) each side of this central value. The span was set for each field site so that adequate smoothing was achieved yet still allowing changes in the shape of the relationship with varying stage to be represented. For all sites except Ebbesbourne Wake, a span of 25 points was found to be acceptable. The span which provided a sufficient degree of smoothing was chosen after testing a range of span sizes. A higher span of 50 points was necessary at Ebbesbourne Wake to smooth small scale fluctuations evident at lower stages. The LOWESS fit then takes each stage-discharge measurement in turn, samples around that central point using the defined span, and fits a weighted linear regression to provide an estimate of discharge at that stage value. This process is then repeated for all of the data points to provide a single LOWESS fit. The estimated variance within each data window for each stage-discharge value was also calculated to represent the scatter in the residual error from the LOWESS fit, which can be used to quantify the uncertainty of the fit. Here, we make the assumption that within the span the residual errors are homoscedastic, independent and randomly distributed. Thus, at each stage value we have an estimated value of discharge (our single LOWESS fit) and variance which describes our 90% uncertainty bounds. This method has advantages over the usual parametric regression techniques as it can more easily account for irregular shapes of the stage-discharge curves and can deal with data which are non-normally distributed, such as the data presented in this paper. The autocorrelation in the discharge error time series was examined and showed that errors could be

represented by random noise over time scales of 2-3 days. Therefore, errors were treated as independent but heteroscedastic in the error model used to sample multiple representative hydrographs and therefore quantify discharge uncertainties through time (see below). The error model was used to generate 100 such time series of errors which were then combined with the original time series. 100 iterations were used throughout this study on the basis that the 10th to 90th percentiles of the final time series distributions were found to be stable and did not change appreciably with additional iterations.

Uncertainty in the daily laboratory nutrient concentration data was calculated using repeated analysis ($n=10$) of a series of laboratory standard solutions. The standard deviation of errors was calculated for each concentration measured and it was found that, for both NO₃-N and TP, the standard deviation increased approximately linearly with concentration magnitude (see section 3.2 for further details). Therefore, a linear model was fitted to the data points ($r^2 = 0.97$ and 0.94 for NO₃-N and TP, respectively) and used to determine the standard deviation of the errors at each concentration in the time series. It was assumed that, for both NO₃-N and TP, the errors were temporally independent as the samples were analysed in the laboratory in batches in a random order. These statistics were then used to model the data error (see below). As with the discharge time series, 100 iterations of the error time series were generated and combined with the original data set.

Uncertainties in the high-frequency in-situ nutrient concentration measurements were determined by comparing the sensor-derived data with the independent daily resolution samples collected via the ISCO autosampler and analysed in the laboratory. The routine approach would be to assume that the laboratory data provides the best estimate of the ‘true’ nutrient value, however in this case the errors derived for the laboratory data were cascaded through and therefore were included in the overall uncertainty estimate of the sensor data. 100 estimates of the sensor data error statistics were calculated from 100 iterations of the possible laboratory nutrient series and the results used to model 10,000 iterations of the errors in the sensor data (see Figure 3). The lag-1 autocorrelation (α) was calculated for the paired sensor-lab series at daily resolution, however α used in the error modelling should relate to the timestep of the sensor series (30 min). Therefore, the lag-1 autocorrelation at a 30 min timestep was calculated using the de-correlation length from the daily timestep, assuming an exponential decay function (Evensen, 2003):

$$\alpha_k = e^{-\frac{\Delta t_k}{\tau}}$$

where, for specific timestep k , Δt is the length of the new timestep (s) and τ is the de-correlation length (s).

2.3.2 Modelling errors

A simple error model was used to generate replicate error time series for the discharge and nutrient data including heteroscedastic and serially-correlated errors as appropriate. Analysis of the sensor data showed that a 1st-order model was adequate to describe the autocorrelated datasets and therefore it was unnecessary to implement a more complex approach. Therefore, a 1st-order autoregressive model was used to generate serially-correlated errors (Evensen, 2003; Garcia-Pintado *et al.*, 2013):

$$q_k = \alpha q_{k-1} + \sqrt{1 - \alpha^2} W_k$$

where q_k is the error at time k , α is temporal correlation at lag 1 and W_k is random white noise at time k . The white noise series was produced with a mean of 0 and a variance of 1. A post-processing step was then performed to scale the white noise component by a time series of standard deviations derived from the uncertainty analysis in the heteroscedastic case (discharge and laboratory nutrient data) and by a single standard deviation in the homoscedastic case (nutrient sensor data).

2.3.3 Nutrient load analyses

The 100 replicate discharge and nutrient time series were paired and used in combination to directly calculate 10,000 estimates of the nutrient load over the monitored period. Gaps exist in the data records at all sites (maximum gap 10 days due to equipment failure) and because of this the daily loads were calculated as kg ha^{-1} based on the number of days of data available in the record and subsequently extrapolated to annual estimates. This approach was adopted rather than using gap-filling methodologies as this would have introduced additional assumptions and uncertainty into these estimates and this was beyond the scope of this paper. We also compared our discharge-nutrient uncertainty load estimation methodology to a) using the discharge uncertainties with the original nutrient time series, and b) using nutrient uncertainties with deterministic discharges. This allowed us to quantify the independent effects of the discharge and nutrient data uncertainties and determine which one dominated uncertainties in the resultant load estimates.

3. Results

3.1 Discharge uncertainty

The changing relationship between stage and discharge, calculated using the velocity-area method, at each site split by season for the two year monitoring period can be seen in Figure 4. All four sub-catchments show marked changes in the stage-discharge curves through time. Without such

knowledge and the usual application of traditional stationary stage-discharge relationships, these seasonal changes would introduce a further epistemic structural error component into our measured discharges. The most defined shifts were seen in the chalk sub-catchments, where there are clear separations between the clusters of points. At Brixton Deverill, for instance, the stage-discharge relationship was relatively stable during the winter and spring, but large shifts can be observed during the summer and autumn months when vegetation growing on the channel banks and from the stream bed causes a decrease in measured velocities at a given stage. For example, at a stage height of 0.7 m the discharge during winter months was between 3.5 and 5 m³ s⁻¹, however when the same stage was reached during the summer and autumn months the calculated discharge was between 1.5 and 2.5 m³ s⁻¹. During autumn, the relationship switches back toward the winter rating due to vegetation die back and hence increased velocities for a given stage. While a similar pattern can be observed at the second chalk site at Ebbesbourne Wake, the shifts are less obvious on the basis of season alone. While there was a general shift towards slower velocities during summer and autumn months, the step changes in the relationship were caused by large flow events producing hysteretic effects, perhaps from scour and channel bed changes. At the clay sites, changes in the stage-discharge relationship were less pronounced, yet at Priors Farm at a stage height of 0.8 m winter and spring flows were between 0.6 and 1.2 m³ s⁻¹ compared with 0.4 and 0.6 m³ s⁻¹ during summer and autumn, respectively. At higher stage values, the discharges were similar in all seasons, possibly because at stages exceeding 0.9 m, the culvert was drowned and the flow was routed out-of-bank reducing the impact of in-stream vegetation on water velocities. In contrast, at Cool's Cottage, higher velocities and therefore higher flows were observed during spring and summer than autumn and winter. This suggests that there was a different or additional control on the stage-discharge relationship at this site rather than just the effect of in-channel vegetation growth.

At each site, a period which exhibited a stable stage-discharge relationship was chosen to estimate the error uncertainty associated with the in-stream measurements as described earlier. Figure 5 shows the selected data for each site along with the LOWESS best fit and 95% confidence intervals. At Brixton Deverill, the data from winter 2012 and spring 2013 were used and Figure 5a shows that the error bounds were small at all stages, although the errors at stages above 0.5 m were larger than those during low flow periods. The standard deviation of the residuals for each of the measured stage heights ranged from ±2.2-9.1%. At the other chalk site, Ebbesbourne Wake (Figure 5b), again the magnitude of the standard deviations across the measured stages was small, ranging from ±0.3-3.2%. At the clay sites, the variance was generally greater, particularly at higher stage heights. The standard deviation of residuals at the clay sites ranged from ±7.8-25.5% and ±2.9-9.1% for Priors Farm and Cool's Cottage, respectively. The standard deviations derived for each stage height recorded were

used according to the method described in section 2.3.1 to model 100 iterations of the discharge time series including errors and these can be seen in Figure 6 (10th- 90th percentiles).

3.2 Nutrient uncertainty

Uncertainty associated with laboratory analysis of NO₃-N showed that the errors were heteroscedastic, with the standard deviation of errors ranging from 10.8 to 1.9% as the concentration increased from 0.5 to 20 mg L⁻¹. A similar pattern was observed for uncertainty in TP determination, with the standard deviation of errors decreasing from 22 to 2.5% for concentrations ranging from 0.01 to 1 mg L⁻¹. These error statistics were used to generate multiple iterations of the NO₃-N and TP data using the error model as described in sections 2.3.1 and 2.3.2. The resulting multiple datasets (10th- 90th percentiles) are shown in Figures 7 and 8.

In contrast to the laboratory data, the sensor data had larger associated measurement uncertainty for both determinands (see Figure 9). These final uncertainty estimates are a combination of the direct measurement uncertainty and the uncertainty associated with the laboratory data used to 'ground-truth' the sensors (see Figure 3). Both NO₃-N and TP were shown to be homoscedastic in their errors, however there was significant temporal correlation which was included in the error estimates produced by the error model (see section 2.3.2). This resulted in a standard deviation of errors of between 0.35 and 0.41 mg L⁻¹ for NO₃-N and 0.098 and 0.11 mg L⁻¹ for TP.

3.3 Impact of uncertainties on nutrient load estimation

Using the non-parametric measurement uncertainties and the error model described previously, resampled datasets could be generated to quantify the overall uncertainties in the time series of data. The results can be seen in the series of boxplots presented in Figures 10, 11 and 12.

Figure 10 shows the ranges of load estimates of NO₃-N at each field site from a) discharge uncertainty, b) nutrient uncertainty and c) both (combined) sources of uncertainty. At the two chalk sites, Brixton Deverill and Ebbesbourne Wake, the discharge and nutrient uncertainty contributes close to equal proportions with the total load uncertainty resulting in possible annual loads between 18.5 and 18.8 kg ha⁻¹ at Brixton Deverill and 58.9 and 60 kg ha⁻¹ at Ebbesbourne Wake. This equates to an average load discrepancy of 1.1% of the median at Brixton Deverill and 1.9% of the median at Ebbesbourne Wake. These discrepancies will have an impact on the estimation of the water pollutant environmental damage costs. In the UK it is suggested by DEFRA that the average cost of excess NO₃-N is £170.87 t⁻¹ compared with £25,690 t⁻¹ for TP (DEFRA, 2012). In terms of estimated environmental damage due to excess NO₃⁻, the total cost at Brixton Deverill varies from £15,782 to £16,038, and at Ebbesbourne Wake from £8,226 to £8,480 (for whole catchments). In contrast, at the clay sites, Priors Farm and

Cool's Cottage, the largest proportion of the uncertainty is derived from the discharge measurements. It is important to note that the median annual $\text{NO}_3\text{-N}$ load estimate changes depending on whether the discharge uncertainty is included or not, shifting from 6.83 to 7.05 kg ha^{-1} at Priors Farm and to a lesser extent from 9.38 to 9.43 kg ha^{-1} at Cool's Cottage when discharge uncertainty is also included. The total range of annual $\text{NO}_3\text{-N}$ loads estimates varied from 6.7 to 7.4 kg ha^{-1} at Priors Farm and from 9.2 to 9.7 kg ha^{-1} at Cool's Cottage. This results in a load discrepancy of 9.9% of the median at Priors Farm and 5.3% of the median at Cool's Cottage. This results in NO_3^- damage costs between £566 and £625 at Priors Farm and between £266 and £280 at Cool's Cottage.

A very similar pattern of results was observed for the TP loads for the four field sites (Figure 11), with discharge and nutrient uncertainty playing almost equal roles in the overall load uncertainty in the chalk sites and the discharge uncertainty being the most important factor at the clay sites. The combined discharge and nutrient uncertainty resulted in a range of annual load estimates from 0.55 to 0.57 kg ha^{-1} at Brixton Deverill, 1.51 to 1.56 kg ha^{-1} at Ebbesbourne Wake, 2.08 to 2.30 kg ha^{-1} at Priors Farm and 0.96 to 1.01 kg ha^{-1} at Cool's Cottage. These ranges result in an overall load discrepancy of 3.6% of the median at Brixton Deverill, 3.3% at Ebbesbourne Wake, 10% at Priors Farm and 5.1% at Cool's Cottage. There is also an implication for the estimated environmental TP damage costs for the catchments, ranging from £70,961 to £73,542 for Brixton Deverill, £32,276 to £33,345 for Ebbesbourne Wake, £26,558 to £29,368 for Priors Farm and £4,193 to £4,411 for Cool's Cottage.

Figure 12 shows the range of loads obtained from the multiple estimates of the discharge time series in combination with the uncertain nutrient estimates from the in-situ sensor data at Brixton Deverill. The results clearly show that for both nitrate and TP the dominant source of uncertainty derives from the sensor measurement errors and the discharge uncertainty in this case plays a minor role. The range of annual load estimates including all uncertainty was 27.5 to 30.1 kg ha^{-1} for $\text{NO}_3\text{-N}$ (damage costs £23,487 - £25,654) and 0.36 to 0.85 kg ha^{-1} for TP (damage costs £45,854 - £110,145). This results in load discrepancies of 9% of the median and 83% of the median for $\text{NO}_3\text{-N}$ and TP, respectively.

4. Discussion

Importance of discharge uncertainties in water quality analyses

The high-frequency monitoring of both stage height and velocity at the four field sites has provided valuable insights into how stage-discharge relationships can vary seasonally and between large discharge events. This level of detail of measurement is often not available in research studies and so uncertainties associated with rating curve construction and parameter measurements are often acknowledged but not fully characterised, particularly for these non-stationary epistemic error

characteristics. The field sites considered in this study all have stable concrete structures, provided by culvert or bridge infrastructure or by a concrete-block cross-section for the purpose of the study. However, these structures have still generated significant discharge uncertainties in these small headwater streams. This finding is supported by evidence presented by Tomkins (2014) who, in a review of the quality of a series of gauging sites, showed that a large proportion of the sites noted as being 'suspect' had an artificial control structure installed. Furthermore these results show seasonal changes in stage-discharge data primarily driven by summer in-channel vegetation growth, which must be taken into account. We have shown this seasonality is more extreme in the chalk streams compared to the catchments underlain by clay. This is most likely due to the large amount of vegetation (macrophyte) growth which commonly occurs in chalk streams during the summer. The dominant macrophyte species of English chalk streams is *Ranunculus* spp. (Clarke *et al.*, 2006), which often occurs in dense stands that can cover up to 80% of a given reach (Cotton *et al.*, 2006; Wharton *et al.*, 2006) and hence modify channel velocities significantly. In contrast, in clay streams the water tends to be more turbid, limiting in-stream macrophyte growth. However, a small seasonal effect is still seen at the clay sites most likely due to the growth of bank or stream margin vegetation close to the monitoring point. At the Cool's Cottage field site, the relationship appears reversed with higher velocities for the same stage height in the summer compared with the winter. Here, other local factors may come into play, such as backwatering, sediment build-up and fallen debris which cannot be easily quantified.

As a result of the non-stationarity of the stage-discharge relationship, it is important to consider the potential for the increased load estimation error associated with the use of a single rating curve. This reflects the situation of increased epistemic uncertainty by not taking these structural changes into account. At Brixton Deverill, a dual system is in place using discharge data provided from the ultrasonic EA gauge, but also an independent manual gauging record prior to 2010. If this rating curve had been used rather than the velocity-area method over this same time period, it would have had a significant impact on the total discharge estimation and hence an impact on the nutrient loads estimated, over and above that characterised with the above methods. Total discharge would have been $3.28 \times 10^7 \text{ m}^3$ (360 mm y^{-1}) compared with $2.3 \times 10^7 \text{ m}^3$ (252 mm y^{-1}) using the velocity-area method. This result is supported by recent work by Birgand *et al.* (2013) who investigated the magnitude of uncertainties associated with the use of stage-discharge curves in small short-term projects. This evidence highlights the importance of using an appropriate technique to monitor discharge as well as accounting for uncertainty within the measurement as it can have a significant impact on the load estimates produced.

Measurement and analytical errors are generally considered to be the smaller components of the total uncertainty associated with load estimation (Rode and Suhr, 2007), compared with uncertainties associated with sampling i.e. frequency and representativeness. However, work presented in this paper shows that significant errors can still exist due to the measurement of velocity and stage height which need to be appropriately accounted for alongside sources of nutrient uncertainty in any load estimation procedure. It is also important to note that the significance of discharge measurement error is likely to be site-specific, yet warrants investigation and quantification. In this paper an error modelling methodology was adopted which allowed the characterisation of the dominant structural and random components of the error, as well as the incorporation of both heteroscedastic and temporally correlated errors where appropriate, overcoming the problem of the common assumption that all uncertainties are random and homoscedastic in their nature (Kulasova *et al.*, 2012; Honti *et al.*, 2013). This new framework for charactering discharge uncertainty generated errors up to approximately $\pm 26\%$ which agrees with estimates made by Pelletier (1988), however these are higher than those presented by Leonard *et al.* (2000), who recorded errors in the range of $\pm 5\text{--}6\%$.

An additional summary of discharge uncertainty ranges from a wide range of studies were provided by (McMillan *et al.*, 2012). Errors for velocity-area techniques were documented in all studies as being up to 20%, again concurring with the results presented herein. Despite the general stage-discharge relationship being more seasonally variable at the chalk sites, the measurement uncertainty was shown to be larger in the clay sites. The chalk sites had errors up to $\pm 9\%$, whereas errors were recorded up to $\pm 26\%$ at the clay sites. A possible explanation for this could be the ability for the ultrasonic velocity probes to operate effectively in streams which have a higher quantity of suspended sediment, which was often observed in the surface water-driven clay sites. Nord *et al.* (2014) recorded an underestimation of velocity of between 10-15% when there were significant quantities of fine sediment ($<150\text{ }\mu\text{m}$) in suspension. Sediment in suspension can reduce velocity probe accuracy, and sites with higher sediment loads will also be more susceptible to sedimentation of the channel cross-section during low flows, leading to further errors associated with the changing stream geometry. It is also important to note that the magnitude of these errors is comparable to those quantified from rating curve studies, even though direct measurements used in the velocity-area method are usually assumed to be more accurate. Discharge calculated via the velocity-area method is often used as the 'true' benchmark from which to test the accuracy of the rating curve methodology, which suggests that studies could be underestimating the overall uncertainty when structural error components are not included.

Nutrient load uncertainty estimates

This study also explored the impact of the discharge uncertainty (combined with nutrient concentration uncertainty) on annual nutrient load estimate uncertainty. The results clearly showed that where high precision laboratory information is available, the dominant source of uncertainty was from the discharge estimation, producing variations in loads of up to 2% at the chalk sites and 12% at the clay sites. However, where the in-situ sensor data were available, the dominant source of uncertainty came from the sensor information, producing loads with a variability of up to 83% in the case of TP and 9% for NO₃-N. It should be noted that the uncertainties were characterised by comparing the sensor values with paired daily laboratory samples with the assumption that these provide our best estimate of the nutrient concentrations at that point in space and time. Given the large number of paired samples available over the two year monitoring period ($n = >600$), this assumption was deemed appropriate. These uncertainties were considerably larger in the case of TP than those produced from the lab data (below 10%) and comparable for the NO₃-N. However, the in-situ chemistry provides the opportunity for much higher frequency sampling which was shown to be key to obtaining accurate loads. The median annual NO₃-N load increased from 18.7 to 28.8 kg ha⁻¹ and the median annual TP load increased from 0.56 to 0.6 kg ha⁻¹ by increasing the sampling frequency from daily to 30 minute resolution. This result poses an interesting question as to which information is more useful, high frequency (and potentially more accurate) but lower precision information which allows flashy storm events to be resolved, or, lower frequency (daily) but higher precision information which may miss some storm events but produces data with lower uncertainties. Given that load estimate uncertainties also increase with decreasing sampling frequency, it is likely that when all sources of uncertainty are brought together, there will be an optimal sampling frequency at which total load estimate uncertainty is minimised. The ideal scenario would be to use the high precision laboratory analysis and apply to more frequent sub-daily sampling regimes; however this is unlikely to be time or financially viable in most cases. Therefore, a judgement has to be taken regarding what compromise should be made and this is likely to depend on the research question being investigated and the sensitivity to an erroneous quantification of loads. It is also important to highlight that the choice of method for dealing with data gaps could have an impact on the absolute values of the loads calculated. In this study the load was calculated based on complete days. The overall load estimates were then scaled up to annual loads by assuming that that mean daily load was representative of the missing data. This could introduce skew into the data if, for example, all of the missing data was during a particular season, for example. However, it was deemed the most appropriate action for this study as the gaps were distributed throughout the year and any method of filling in missing data was likely to introduce additional errors into the final analysis.

The impact of the velocity-area method versus a rating curve at Brixton Deverill on load estimation was also tested and the results showed that the TP load over the two year monitoring period would have been over-estimated by 1 tonne (0.2 kg ha^{-1}) and the $\text{NO}_3\text{-N}$ load by 67 tonnes (13.3 kg ha^{-1}) if the rating curve approach had been used. This uncertainty is larger than any of the measurement uncertainties reported in this paper (TP 22% and $\text{NO}_3\text{-N}$ 39%), emphasising the importance of effective characterisation of flow regimes in small-headwater catchments. There is currently a lack of literature to compare the values obtained here against, as little work has been done so far to investigate the impact of discharge uncertainty on nutrient load estimates. The bulk of the current literature discusses uncertainties in load estimation with regards to sampling frequency of the nutrient data (Johnes, 2007; Moatar and Meybeck, 2007; Jones *et al.*, 2012; Worrall *et al.*, 2013) and also load estimation method or model (Johnes, 2007; Verma *et al.*, 2012; Raymond *et al.*, 2013). These are additional sources of uncertainty which were not considered in this paper due to the relatively high frequency and paired nature of the data available and therefore the use of a load model was not necessary in this instance. However, it is important to recognise that for studies where a load model is utilised any uncertainty incorporated from the load model or estimation method would be in addition to the sources of uncertainty discussed here derived from the discharge and nutrient measurements.

With this in mind, it is vital that sources of error are adequately characterised. Even in scenarios where the errors are thought to be minimal they could become significant during further post-processing of the data. These findings have several important implications for the water quality community. First, realistic estimation of nutrient loads is critical for accurate assessment of current water quality status as well as determining whether the system is in a stationary state. Second, it has implications for our ability to assess the effectiveness of land management strategies which have been implemented to mitigate diffuse agricultural water pollution, especially if changes smaller than the uncertainty ranges are expected or monitored. Third, it has important implications for a range of stakeholders from farmers through to water companies in the context of cost-benefit analysis. For instance, the estimated cost of excess nitrate increases from around £16k to £24k, and for TP from around £72k to £76k when sensor data is used over the laboratory data. However, the uncertainty within the TP sensor data means that there is a large range of possible costs (£42k-£110k) at the catchment scale which, if not accounted for, could have implications for management decisions.

5. Summary and conclusions

The main aim of this study was to assess the magnitude and impact of discharge and nutrient uncertainty on nutrient load estimation through a comprehensive uncertainty analysis approach. Two years of high frequency (15 min) velocity and stage height data were used to examine the

measurement uncertainty associated with discharge. This was then used in conjunction with $\text{NO}_3\text{-N}$ and TP nutrient concentration uncertainty estimates determined from high resolution lab-in-the-field sensor technologies and laboratory based methods to examine variability in annual nutrient load calculations. A framework for estimating uncertainty was used which combined a flexible non-parametric approach to characterising discharge uncertainty, with error modelling that allowed the incorporation of errors which were heteroscedastic and temporally correlated. Analysis of stage and discharge across four field sites which encompassed both chalk and clay streams showed that there were large seasonal fluctuations in the relationship and that it was critical to account for local conditions. In general, it was found that the streams were more efficient at transporting water at a given stage height during the winter and spring months due to the reduced macrophyte growth outside the growing season. This was particularly pronounced at the chalk sites. As a result, Brixton Deverill results showed that the use of a single rating curve would have resulted in a substantial over-estimation of the flow over the monitoring period, by almost 50%. This suggests that when conducting research in small headwater catchments, which are likely to be very dynamic in their behaviour, the use of multiple seasonal rating curves or quasi-continuous monitoring of stage height and velocity is advisable where possible. This study also showed that even using a velocity-area method at sites which had fixed geometries there were still significant uncertainties in the flow estimation, especially in the clay sites (up to $\pm 26\%$) where the water is likely to be more turbid and therefore can hinder the performance of the in-situ sensors.

These errors were then cascaded through to the nutrient load estimation along with errors associated with the measurement of $\text{NO}_3\text{-N}$ and TP in both the field and the laboratory. Where laboratory procedures were used to determine nutrient concentrations, the discharge uncertainty played the most significant role in generating nutrient load estimate uncertainties. However, this was reversed when in-situ high frequency sensor data were examined. Overall the variability in nutrient load estimates ranged from 1.1-9.9% for $\text{NO}_3\text{-N}$ and from 3.3-10% for TP when daily laboratory data were used, rising to a maximum of 9% for $\text{NO}_3\text{-N}$ and 83% for TP when the sensor data were used instead. This poses an interesting question: is it more beneficial to have high frequency, lower precision data or lower frequency but higher precision data streams to estimate nutrient flux behaviours and load estimates in catchments. The answer is likely to depend on the questions being posed.

Overall, this study has provided understanding and quantification of observational uncertainties in discharge and nutrient parameters. The data have shown that even in scenarios which are generally considered to be the 'truth' in many studies (high frequency, paired data) significant uncertainties can exist. The magnitude of the uncertainties has been shown to be catchment dependent and influenced

by local conditions and therefore it is critical to adequately explore these when starting work at a new research site. It is impossible to eradicate these uncertainties; they are a fact of quantifying natural system behaviour. The methodology presented here could be expanded to investigate the impact of monitoring at a fixed point in the context of cross-section variations in determinands. Given an appropriate framework to characterise and account for observational uncertainties, such as the one presented here which allow the quantification of uncertainties from different sources (discharge and nutrients), it will be possible to generate more robust analyses. In turn this will result in better estimation of nutrient fluxes from land to stream, and thereby underpin the development of more reliable land management decisions.

Acknowledgements

The funding provided by Defra project WQ0211 is gratefully acknowledged. Jeff Short, Tony Fox, Carla Richmond, Fiona Grant and Robin Hodgkinson (ADAS) installed and maintained the monitoring stations used to generate the example data discussed in this paper. Geoff Warren, Moragh Stirling and Chris Yates (University of Reading) undertook the laboratory analyses. The permission to install the monitoring equipment generously granted by landowners is acknowledged.

576 6. References

- 577 Allen DJ, Darling WG, Davies J, Newell AJ, Gooddy DC, Collins AL. 2014. Groundwater conceptual
578 models: implications for evaluating diffuse pollution mitigation measures. *Quarterly Journal*
579 *of Engineering Geology and Hydrogeology*, **47**: 65-80. DOI: 10.1144/qjegh2013-043.
- 580 Birgand F, Lellouche G, Appelboom TW. 2013. Measuring flow in non-ideal conditions for short-term
581 projects: Uncertainties associated with the use of stage-discharge rating curves. *Journal of*
582 *Hydrology*, **503**: 186-195. DOI: 10.1016/j.jhydrol.2013.09.007.
- 583 Clarke S, Wharton G, Cotton J. 2006. Spatial and Temporal Variations in the Sediment Habitat of
584 *Ranunculus* spp. in Lowland Chalk Streams – Implications for Ecological Status? *Water Air Soil*
585 *Pollut: Focus*, **6**: 393-401. DOI: 10.1007/s11267-006-9051-4.
- 586 Cotton JA, Wharton G, Bass JAB, Heppell CM, Wotton RS. 2006. The effects of seasonal changes to in-
587 stream vegetation cover on patterns of flow and accumulation of sediment. *Geomorphology*,
588 **77**: 320-334. DOI: 10.1016/j.geomorph.2006.01.010.
- 589 DEFRA. 2012. Estimated damage costs provided to this study by departmental economists.
- 590 Di Baldassarre G, Montanari A. 2009. Uncertainty in river discharge observations: a quantitative
591 analysis. *Hydrology and Earth System Sciences*, **13**: 913-921.
- 592 Evensen G. 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation.
593 *Ocean Dynamics*, **53**: 343-367. DOI: 10.1007/s10236-003-0036-9.
- 594 Garcia-Pintado J, Neal JC, Mason DC, Dance SL, Bates PD. 2013. Scheduling satellite-based SAR
595 acquisition for sequential assimilation of water level observations into flood modelling.
596 *Journal of Hydrology*, **495**: 252-266. DOI: 10.1016/j.jhydrol.2013.03.050.
- 597 Halliday SJ, Wade AJ, Skeffington RA, Neal C, Reynolds B, Rowland P, Neal M, Norris D. 2012. An
598 analysis of long-term trends, seasonality and short-term dynamics in water quality data from
599 Plynlimon, Wales. *Science of the Total Environment*, **434**: 186-200. DOI:
600 10.1016/j.scitotenv.2011.10.052.
- 601 Hamilton AS, Moore RD. 2012. Quantifying Uncertainty in Streamflow Records. *Canadian Water*
602 *Resources Journal*, **37**: 3-21. DOI: 10.4296/cwrj3701865.
- 603 Honti M, Stamm C, Reichert P. 2013. Integrated uncertainty assessment of discharge predictions with
604 a statistical error model. *Water Resources Research*, **49**: 4866-4884. DOI:
605 10.1002/wrcr.20374.
- 606 Horowitz AJ. 2003. An evaluation of sediment rating curves for estimating suspended sediment
607 concentrations for subsequent flux calculations. *Hydrological Processes*, **17**: 3387-3409. DOI:
608 10.1002/hyp.1299.
- 609 Horowitz AJ. 2008. Determining annual suspended sediment and sediment-associated trace element
610 and nutrient fluxes. *Science of the Total Environment*, **400**: 315-343. DOI:
611 10.1016/j.scitotenv.2008.04.022.
- 612 ISO 1088. 2007. Hydrometry -- Velocity-area methods using current-meters -- Collection and
613 processing of data for determination of uncertainties in flow measurement. In: ISO standards.
- 614 ISO 1100-2. 2010. Hydrometry -- Measurement of liquid flow in open channels -- Part 2: Determination
615 of the stage-discharge relationship. In: ISO Standards.
- 616 Jalbert J, Mathevet T, Favre A-C. 2011. Temporal uncertainty estimation of discharges from rating
617 curves using a variographic analysis. *Journal of Hydrology*, **397**: 83-92. DOI:
618 10.1016/j.jhydrol.2010.11.031.
- 619 Johnes PJ. 2007. Uncertainties in annual riverine phosphorus load estimation: Impact of load
620 estimation methodology, sampling frequency, baseflow index and catchment population
621 density. *Journal of Hydrology*, **332**: 241-258. DOI: 10.1016/j.jhydrol.2006.07.006.
- 622 Johnes PJ, Heathwaite AL. 1992. A procedure for the simultaneous determination of total nitrogen
623 and total phosphorus in fresh-water samples using persulfate microwave digestion. *Water*
624 *Research*, **26**: 1281-1287. DOI: 10.1016/0043-1354(92)90122-k.
- 625 Jones AS, Horsburgh JS, Mesner NO, Ryel RJ, Stevens DK. 2012. Influence of Sampling Frequency on
626 Estimation of Annual Total Phosphorus and Total Suspended Solids Loads. *Journal of the*

- American Water Resources Association, **48**: 1258-1275. DOI: 10.1111/j.1752-1688.2012.00684.x.
- Jordan P, Cassidy R. 2011. Technical Note: Assessing a 24/7 solution for monitoring water quality loads in small river catchments. *Hydrology and Earth System Sciences*, **15**: 3093-3100. DOI: 10.5194/hess-15-3093-2011.
- Juston J, Jansson P-E, Gustafsson D. 2014. Rating curve uncertainty and change detection in discharge time series: case study with 44-year historic data from the Nyangores River, Kenya. *Hydrological Processes*, **28**: 2509-2523. DOI: 10.1002/hyp.9786.
- Kronvang B, Bruhn AJ. 1996. Choice of sampling strategy and estimation method for calculating nitrogen and phosphorus transport in small lowland streams. *Hydrological Processes*, **10**: 1483-1501. DOI: 10.1002/(sici)1099-1085(199611)10:11<1483::aid-hyp386>3.0.co;2-y.
- Krueger T, Quinton JN, Freer J, Macleod CJA, Bilotta GS, Brazier RE, Butler P, Haygarth PM. 2009. Uncertainties in Data and Models to Describe Event Dynamics of Agricultural Sediment and Phosphorus Transfer. *Journal of Environmental Quality*, **38**: 1137-1148. DOI: 10.2134/jeq2008.0179.
- Krueger T, Quinton JN, Freer J, Macleod CJA, Bilotta GS, Brazier RE, Hawkins JMB, Haygarth PM. 2012. Comparing empirical models for sediment and phosphorus transfer from soils to water at field and catchment scale under data uncertainty. *European Journal of Soil Science*, **63**: 211-223. DOI: 10.1111/j.1365-2389.2011.01419.x.
- Kulasova A, Smith PJ, Beven KJ, Blazkova SD, Hlavacek J. 2012. A method of computing uncertain nitrogen and phosphorus loads in a small stream from an agricultural catchment using continuous monitoring data. *Journal of Hydrology*, **458**: 1-8. DOI: 10.1016/j.jhydrol.2012.05.060.
- Leonard J, Mietton M, Najib H, Gourbesville P. 2000. Rating curve modelling with Manning's equation to manage instability and improve extrapolation. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, **45**: 739-750. DOI: 10.1080/02626660009492374.
- Liu Y, Freer J, Beven K, Matgen P. 2009. Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error. *Journal of Hydrology*, **367**: 93-103. DOI: 10.1016/j.jhydrol.2009.01.016.
- McGonigle DF, Burke SP, Collins AL, Gartner R, Haft MR, Harris RC, Haygarth PM, Hedges MC, Hiscock KM, Lovett AA. 2014. Developing Demonstration Test Catchments as a platform for transdisciplinary land management research in England and Wales. *Environmental Science: Processes & Impacts*. DOI: 10.1039/C3EM00658A.
- McMillan H, Freer J, Pappenberger F, Krueger T, Clark M. 2010. Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes*, **24**: 1270-1284. DOI: 10.1002/hyp.7587.
- McMillan H, Krueger T, Freer J. 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. *Hydrological Processes*, **26**: 4078-4111. DOI: 10.1002/hyp.9384.
- Moatar F, Meybeck M. 2007. Riverine fluxes of pollutants: Towards predictions of uncertainties by flux duration indicators. *Comptes Rendus Geoscience*, **339**: 367-382. DOI: 10.1016/j.crte.2007.05.001.
- Moatar F, Meybeck M, Raymond S, Birgand F, Curie F. 2013. River flux uncertainties predicted by hydrological variability and riverine material behaviour. *Hydrological Processes*, **27**: 3535-3546. DOI: 10.1002/hyp.9464.
- NIVUS. 2013. KDA Doppler Wedge Sensor: Wedge-shaped flow sensor for channels datasheet.
- Nord G, Gallart F, Gratiot N, Soler M, Reid I, Vachtman D, Latron J, Pedro Martin-Vide J, Laronne JB. 2014. Applicability of acoustic Doppler devices for flow velocity measurements and discharge estimation in flows with sediment transport. *Journal of Hydrology*, **509**: 504-518. DOI: 10.1016/j.jhydrol.2013.11.020.

- Outram FN, Lloyd CEM, Jonczyk J, Benskin CMH, Grant F, Perks MT, Deasy C, Burke SP, Collins AL, Freer J, Haygarth PM, Hiscock KM, Johnes PJ, Lovett AL. 2014. High-frequency monitoring of nitrogen and phosphorus response in three rural catchments to the end of the 2011–2012 drought in England. *Hydrol. Earth Syst. Sci.*, **18**: 3429-3448. DOI: 10.5194/hess-18-3429-2014.
- Pelletier PM. 1988. Uncertainties in the single determination of river discharge: a literature review. *Canadian Journal of Civil Engineering*, **15**: 834-850. DOI: 10.1139/l88-109.
- Rantz SE, al. e. 1982. Measurement and computation of streamflow. In: Water Supply Paper.
- Raymond S, Moatar F, Meybeck M, Bustillo V. 2013. Choosing methods for estimating dissolved and particulate riverine fluxes from monthly sampling. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, **58**: 1326-1339. DOI: 10.1080/02626667.2013.814915.
- Rekolainen S, Posch M, Kamari J, Ekholm P. 1991. Evaluation of the accuracy and precision of annual phosphorus load estimates from 2 agricultural basins in finland. *Journal of Hydrology*, **128**: 237-255. DOI: 10.1016/0022-1694(91)90140-d.
- Robson A, Reed D. 1999. Flood estimation handbook - FEH CD-ROM 3. Institute of Hydrology.
- Rode M, Suhr U. 2007. Uncertainties in selected river water quality data. *Hydrology and Earth System Sciences*, **11**: 863-874.
- Shao QX, Lerat J, Podger G, Dutta D. 2014. Uncertainty estimation with bias-correction for flow series based on rating curve. *Journal of Hydrology*, **510**: 137-152. DOI: 10.1016/j.jhydrol.2013.12.025.
- Tomkins KM. 2014. Uncertainty in streamflow rating curves: methods, controls and consequences. *Hydrological Processes*, **28**: 464-481. DOI: 10.1002/hyp.9567.
- Verma S, Markus M, Cooke RA. 2012. Development of error correction techniques for nitrate-N load estimation methods. *Journal of Hydrology*, **432**: 12-25. DOI: 10.1016/j.jhydrol.2012.02.011.
- Webb BW, Phillips JM, Walling DE. 2000. A new approach to deriving 'best-estimate' chemical fluxes for rivers draining the LOIS study area. *Science of the Total Environment*, **251**: 45-54. DOI: 10.1016/s0048-9697(00)00413-7.
- Westerberg I, Guerrero JL, Seibert J, Beven KJ, Halldin S. 2011. Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. *Hydrological Processes*, **25**: 603-613. DOI: 10.1002/hyp.7848.
- Wharton G, Cotton JA, Wotton RS, Bass JAB, Heppell CM, Trimmer M, Sanders IA, Warren LL. 2006. Macrophytes and suspension-feeding invertebrates modify flows and fine sediments in the Frome and Piddle catchments, Dorset (UK). *Journal of Hydrology*, **330**: 171-184. DOI: 10.1016/j.jhydrol.2006.04.034.
- Worrall F, Howden NJK, Burt TP. 2013. Assessment of sample frequency bias and precision in fluvial flux calculations - An improved low bias estimation method. *Journal of Hydrology*, **503**: 101-110. DOI: 10.1016/j.jhydrol.2013.08.048.
- Yates CA, Johnes PJ. 2013. Nitrogen speciation and phosphorus fractionation dynamics in a lowland Chalk catchment. *Science of the Total Environment*, **444**: 466-479. DOI: 10.1016/j.scitotenv.2012.12.002.

718 **List of Tables**

719 Table 1: Summary characteristics of the four studies field sites.

720

List of figures

Figure 1: Map showing the location of the Hampshire Avon catchment and the location of the Wylfe, Sem and Ebble sub-catchments. Sampling stations at the sub-catchment outlets are shown as black dots.

Figure 2: Photos showing gauging sites on the chalk streams at Brixton Deverill and Ebbesbourne Wake (a, b) and on the clay streams at Priors Farm and Cool's Cottage (c, d).

Figure 3: Flow diagram illustrating uncertainty analysis and error modelling procedures.

Figure 4: Plots showing relationship between stage height and discharge split by season for a) Brixton Deverill, b) Ebbesbourne Wake, c) Priors Farm and d) Cool's Cottage.

Figure 5: Plots showing sub-sets of the stage-discharge data from which uncertainty bounds were quantified. Lowess best-fits are shown by the solid lines and 95% confidence intervals are shown by the dashed lines.

Figure 6: Plots showing discharge measurements and green area represents uncertainty based on the error modelling (10th – 90th percentiles).

Figure 7: Plots showing NO₃-N concentrations from lab analysis of daily samples, where green areas represent the measurement uncertainty (10th- 90th percentiles).

Figure 8: Plots showing total phosphorus concentrations from lab analysis of daily samples, where the green area shows the measurement uncertainty (10th- 90th percentiles).

Figure 9: Plots showing a) NO₃-N and b) total phosphorus concentrations at Brixton Deverill from 30 min resolution in situ sensor data, where the green area shows the measurement uncertainty (10th- 90th percentiles).

Figure 10: Boxplots showing the range of NO₃-N loads including discharge uncertainty, nutrient uncertainty and total uncertainty for a) Brixton Deverill, b) Ebbesbourne Wake, c) Priors Farm and d) Cool's Cottage.

Figure 11: Boxplots showing the range of total phosphorus loads including discharge uncertainty, nutrient uncertainty and total uncertainty for a) Brixton Deverill, b) Ebbesbourne Wake, c) Priors Farm and d) Cool's Cottage.

Figure 12: Boxplots showing the range of a) NO₃-N and b) total phosphorus loads at Brixton Deverill using 30 min resolution sensor data, including discharge uncertainty, nutrient uncertainty and total uncertainty.

753 Table 1: Summary characteristics of the four studies field sites.

River	Sem	Sem	Wylfe	Ebble
Site name	Cool's Cottage	Priors Farm	Brixton Deverill	Ebbesbourne Wake
National Grid reference	ST901297	ST891284	ST858381	ST986243
Catchment area (km ²)	1.70	4.97	50.22	8.32
Elevation (m.a.s.l.) ^a	163	126	189	165
Annual average rainfall (mm) ^{a+}	897	863	967	912
Baseflow Index (BFI) ^a	0.49	0.23	0.93	0.97
Dominant geology	Clay	Clay	Chalk	Chalk
Land use	Lowland livestock grazing, dairy	Lowland livestock grazing, dairy	Sheep grazing, cereal cropping	Sheep grazing, cereal cropping

754 ^a From (Robson and Reed, 1999)

755 ⁺ Average 1961-1990

756

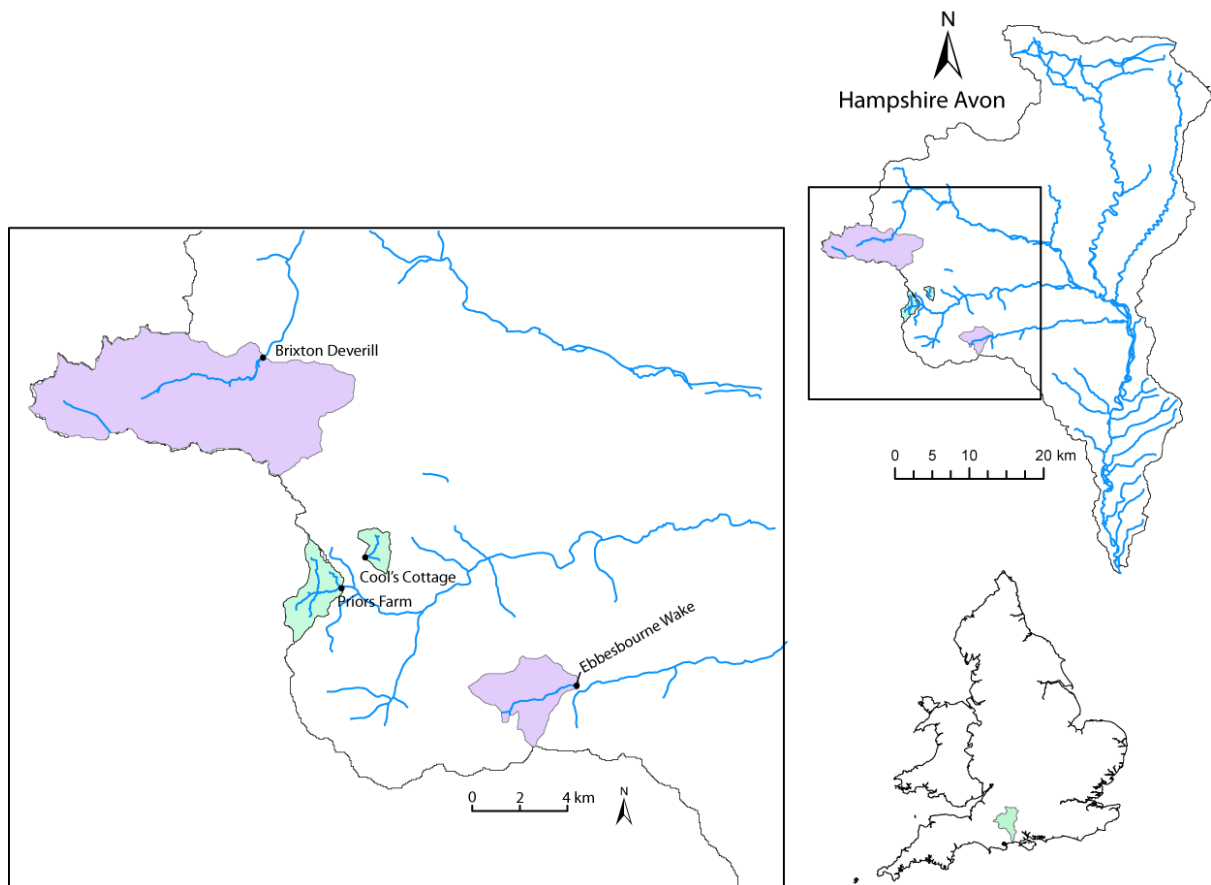


Figure 1: Map showing the location of the Hampshire Avon catchment and the location of the Wylfe, Sem and Ebble sub-catchments. Sampling stations at the sub-catchment outlets are shown as black dots.



Figure 2: Photos showing gauging sites on the chalk streams at Brixton Deverill and Ebbesbourne Wake (a, b) and on the clay streams at Priors Farm and Cool's Cottage (c, d).

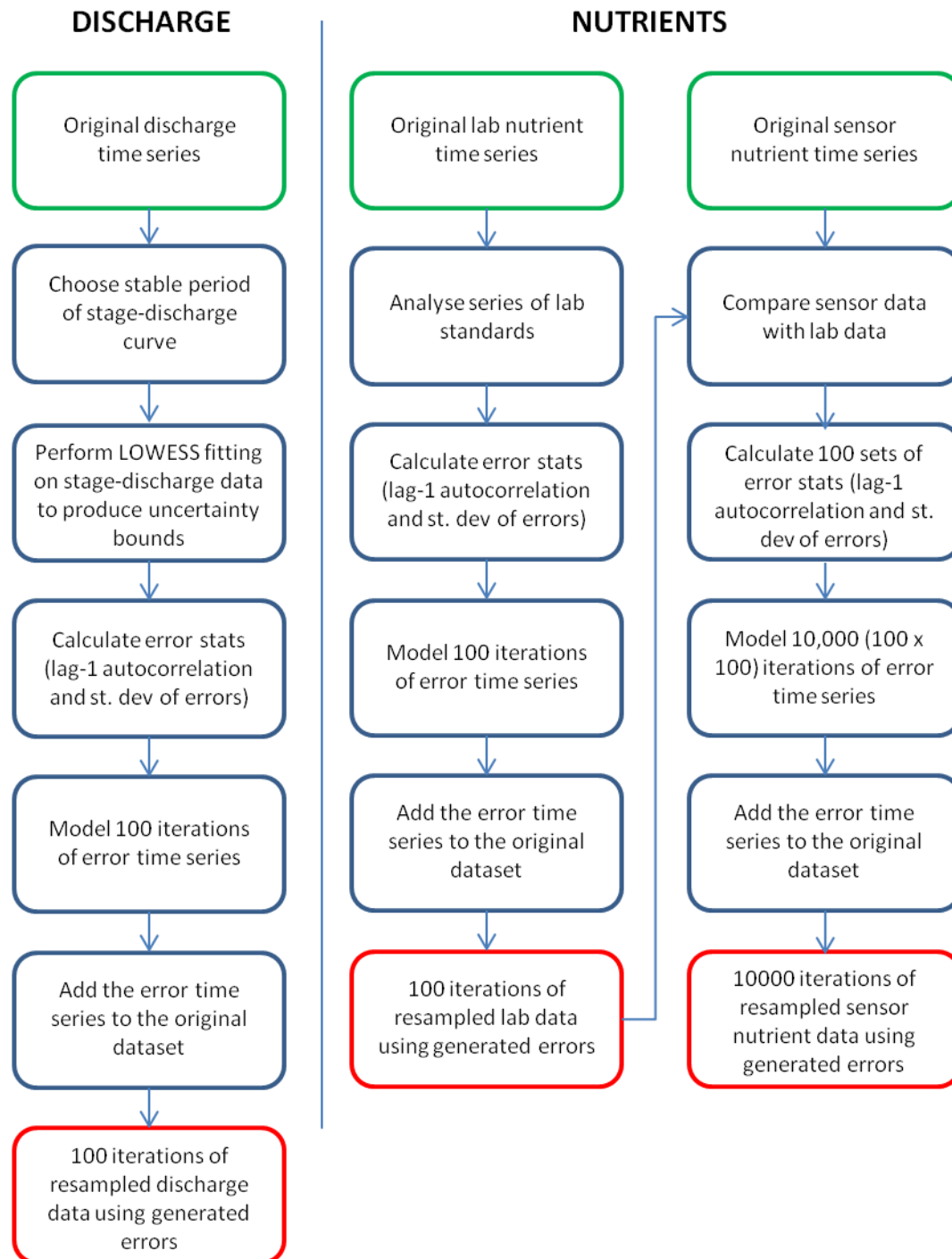


Figure 3: Flow diagram illustrating uncertainty analysis and error modelling procedures.

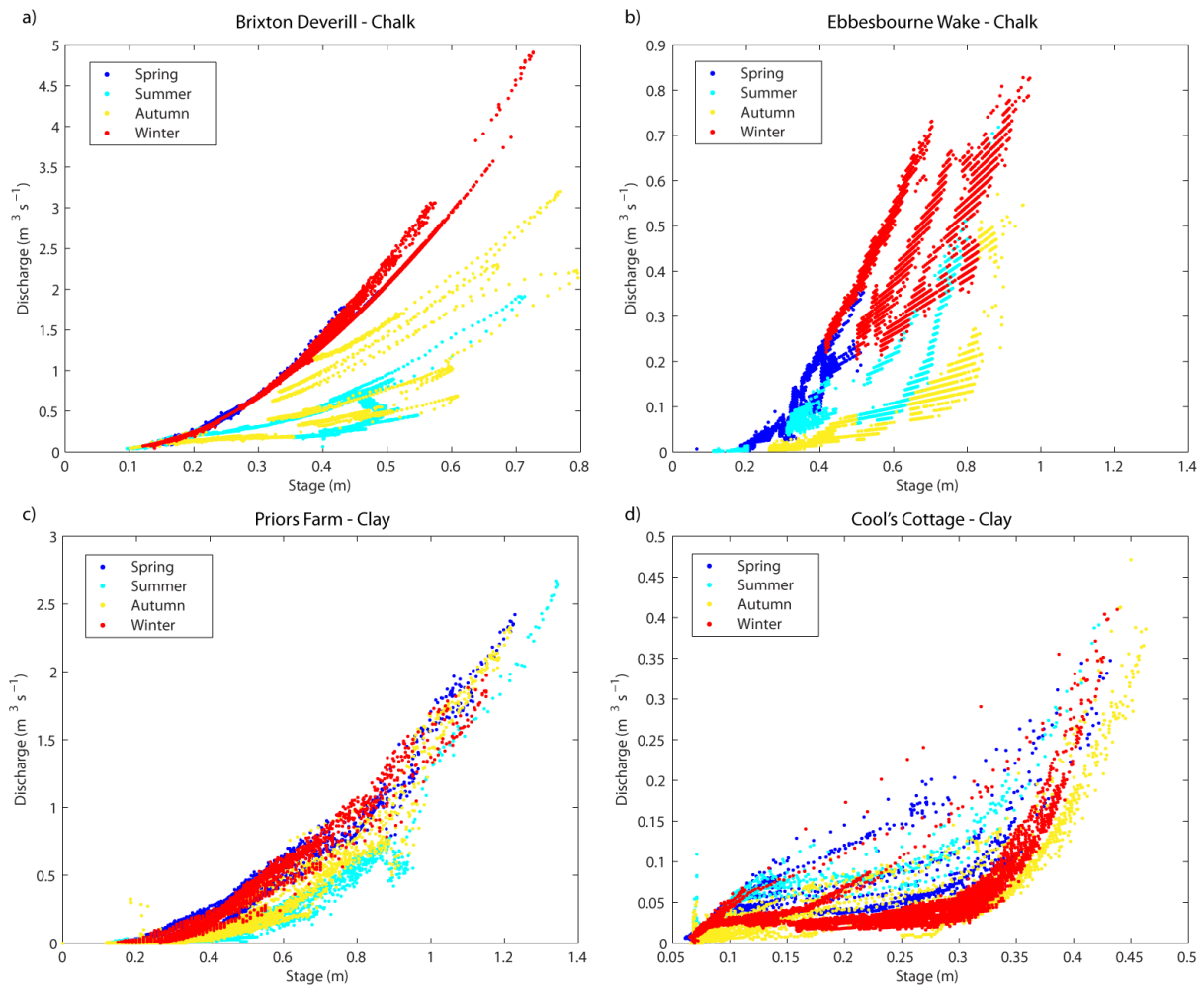


Figure 4: Plots showing relationship between stage height and discharge split by season for a) Brixton Deverill, b) Ebbesbourne Wake, c) Priors Farm and d) Cool's Cottage.

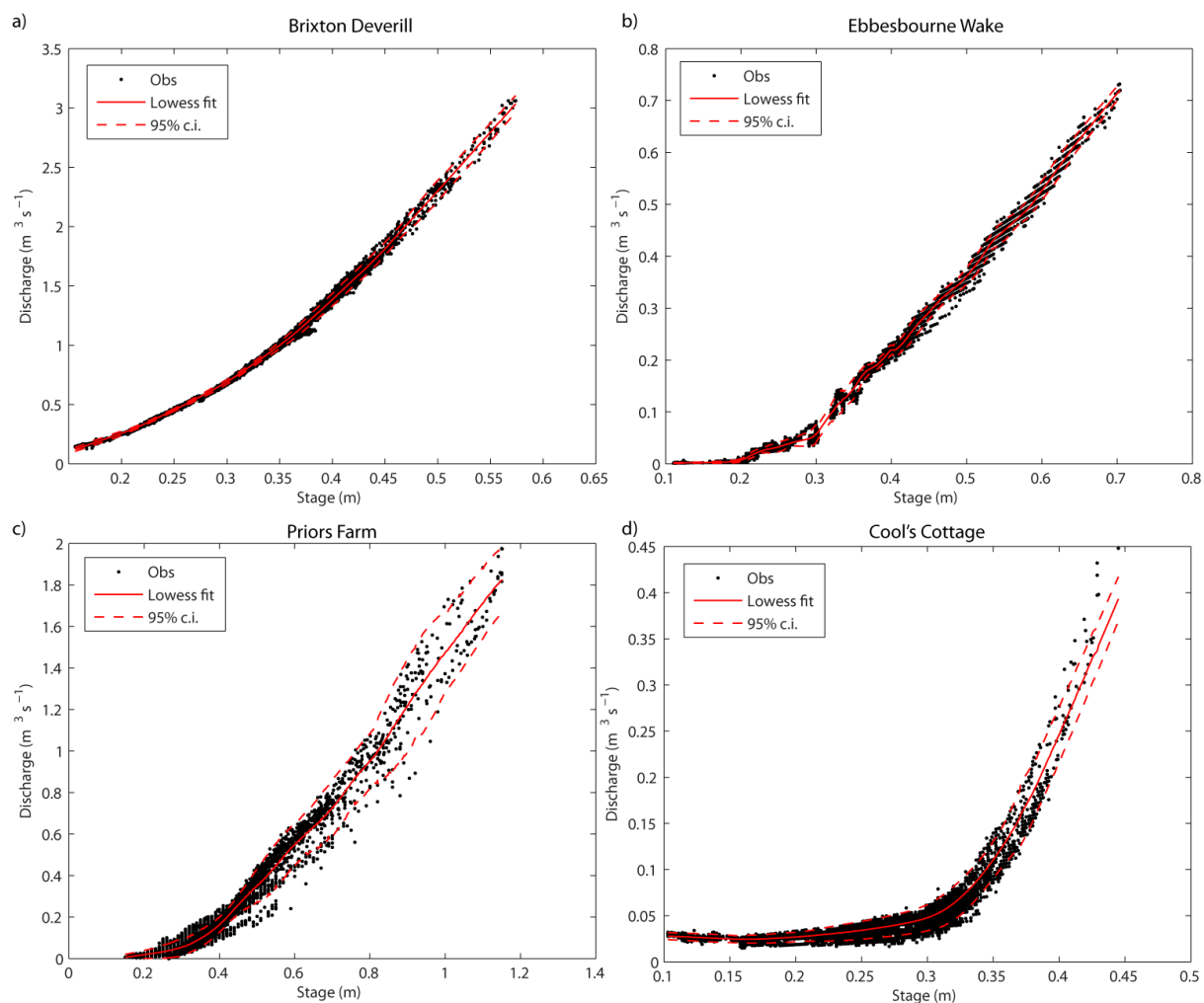
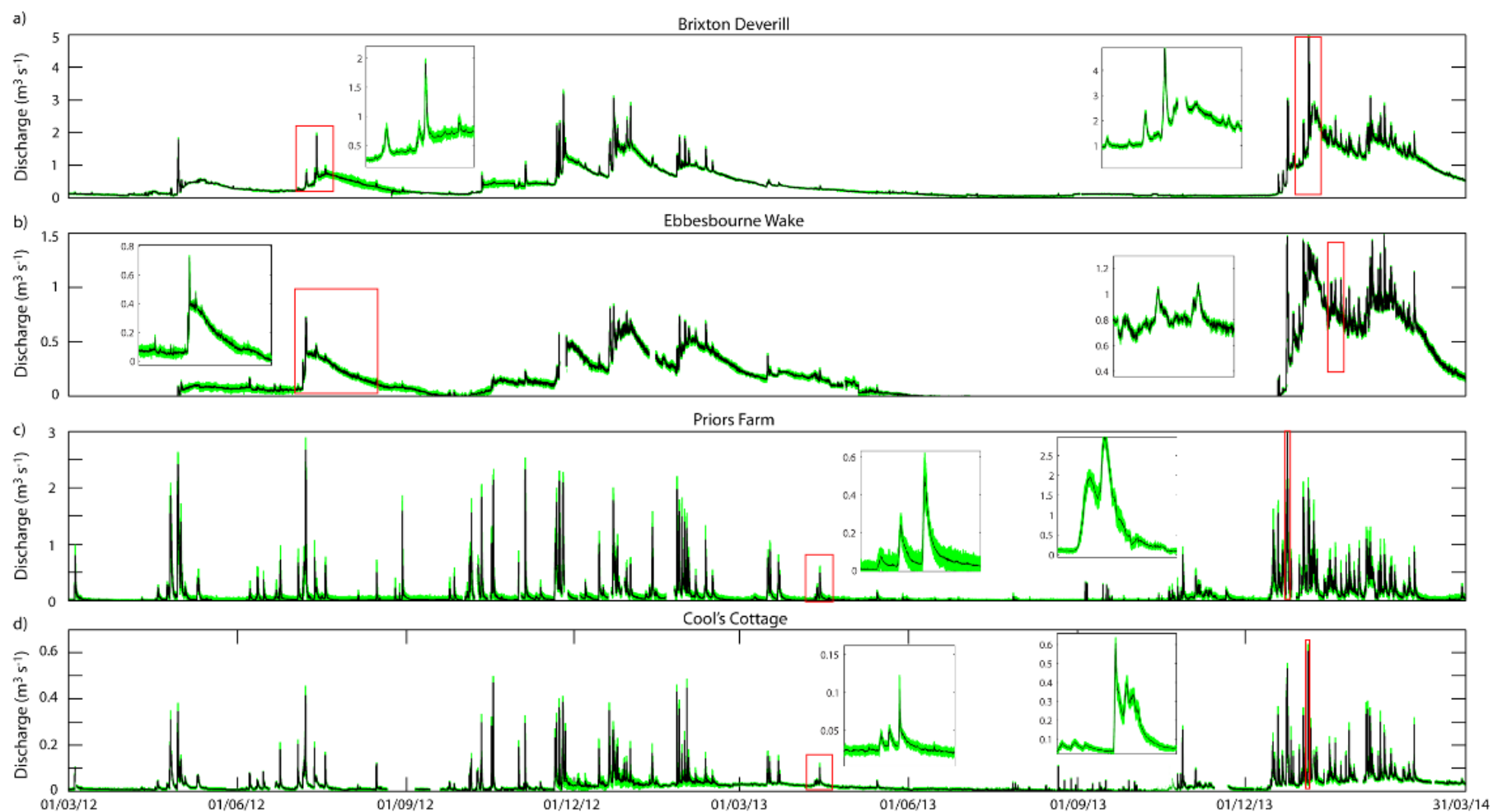


Figure 5: Plots showing sub-sets of the stage-discharge data from which uncertainty bounds were quantified. Lowess best-fits are shown by the solid lines and 95% confidence intervals are shown by the dashed lines.



778

779 Figure 6: Plots showing discharge measurements and green area represents uncertainty based on the error modelling (10th – 90th percentiles).

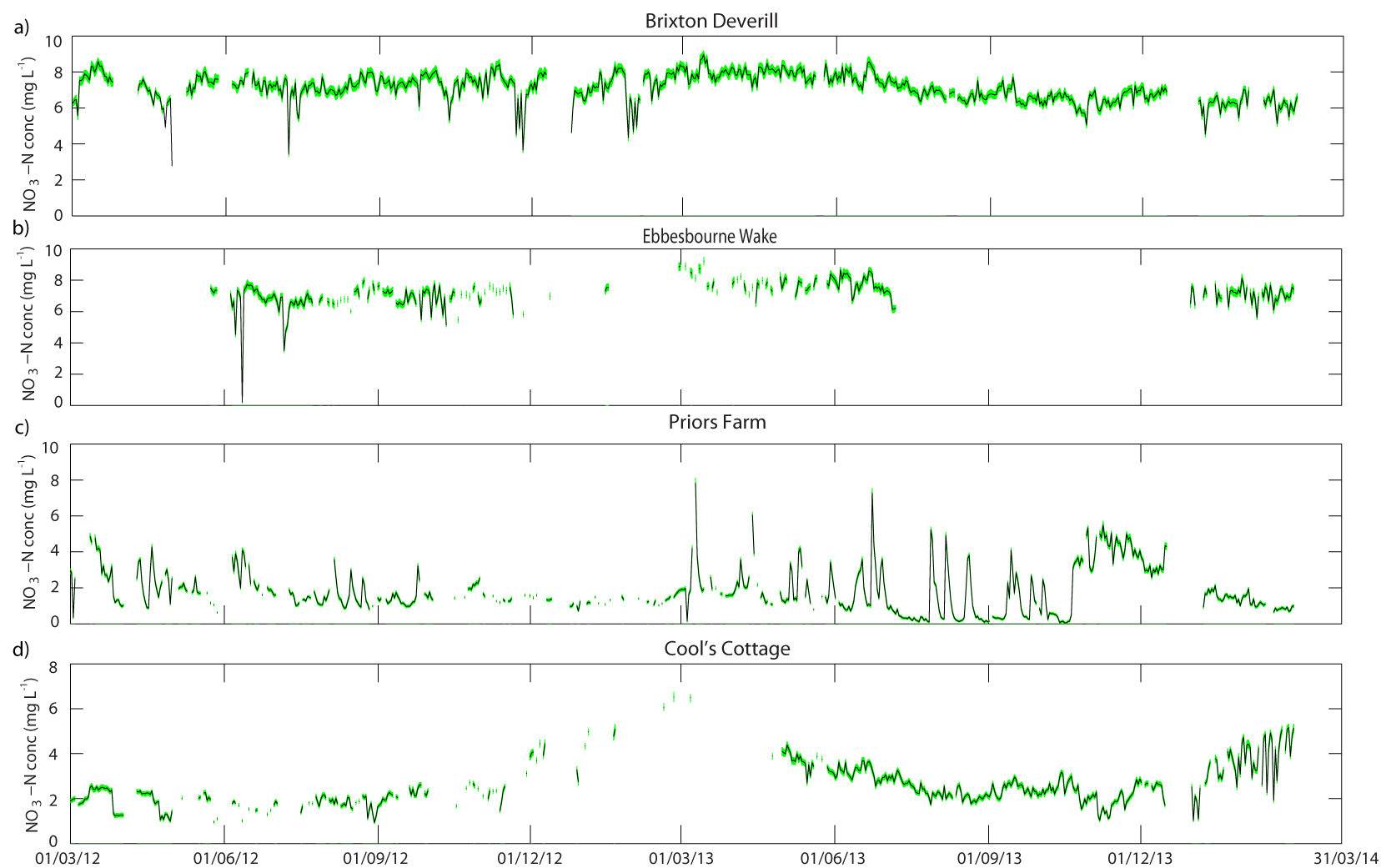


Figure 7: Plots showing $\text{NO}_3\text{-N}$ concentrations from lab analysis of daily samples, where green areas represent the measurement uncertainty (10th- 90th percentiles).

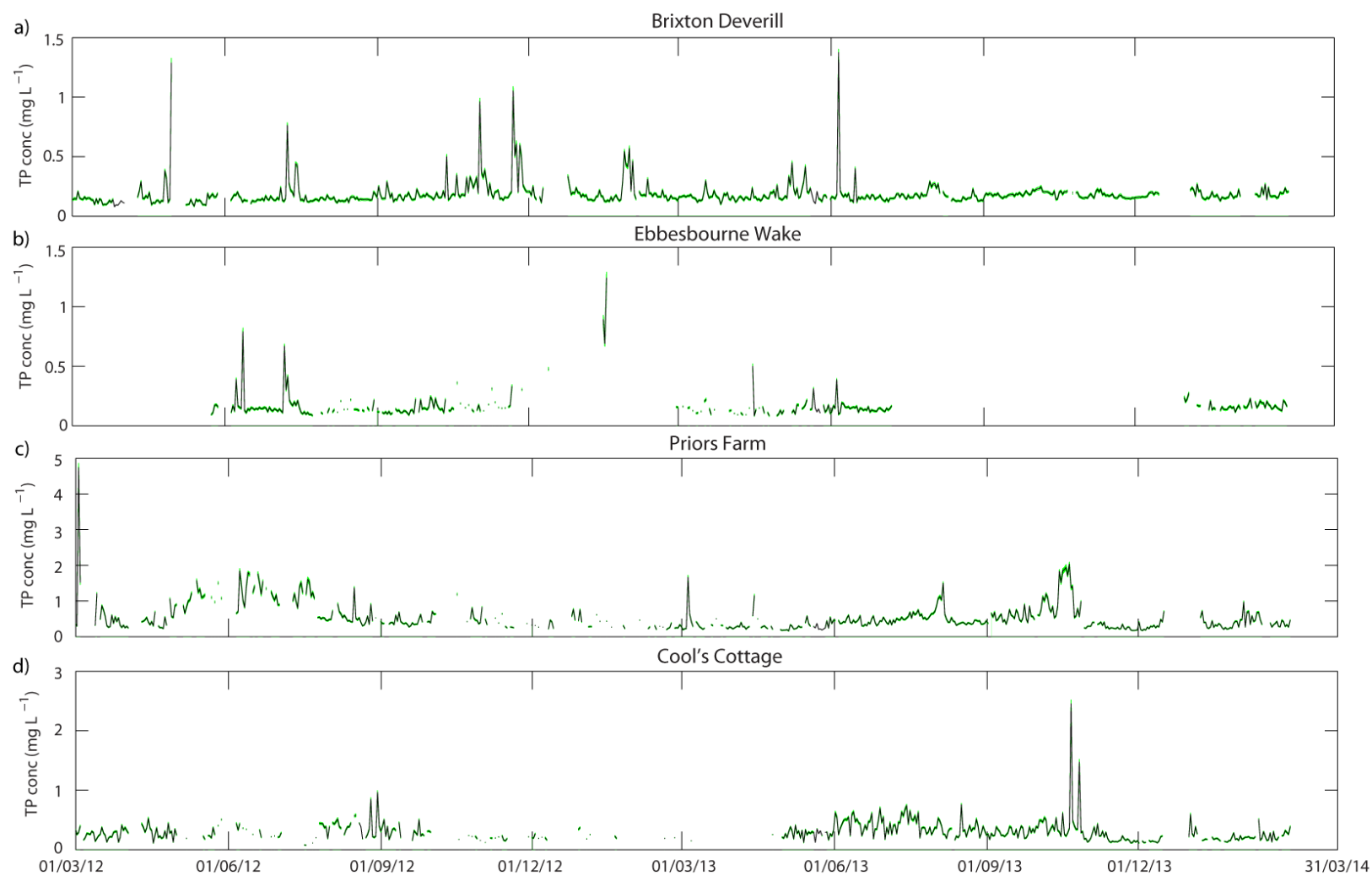
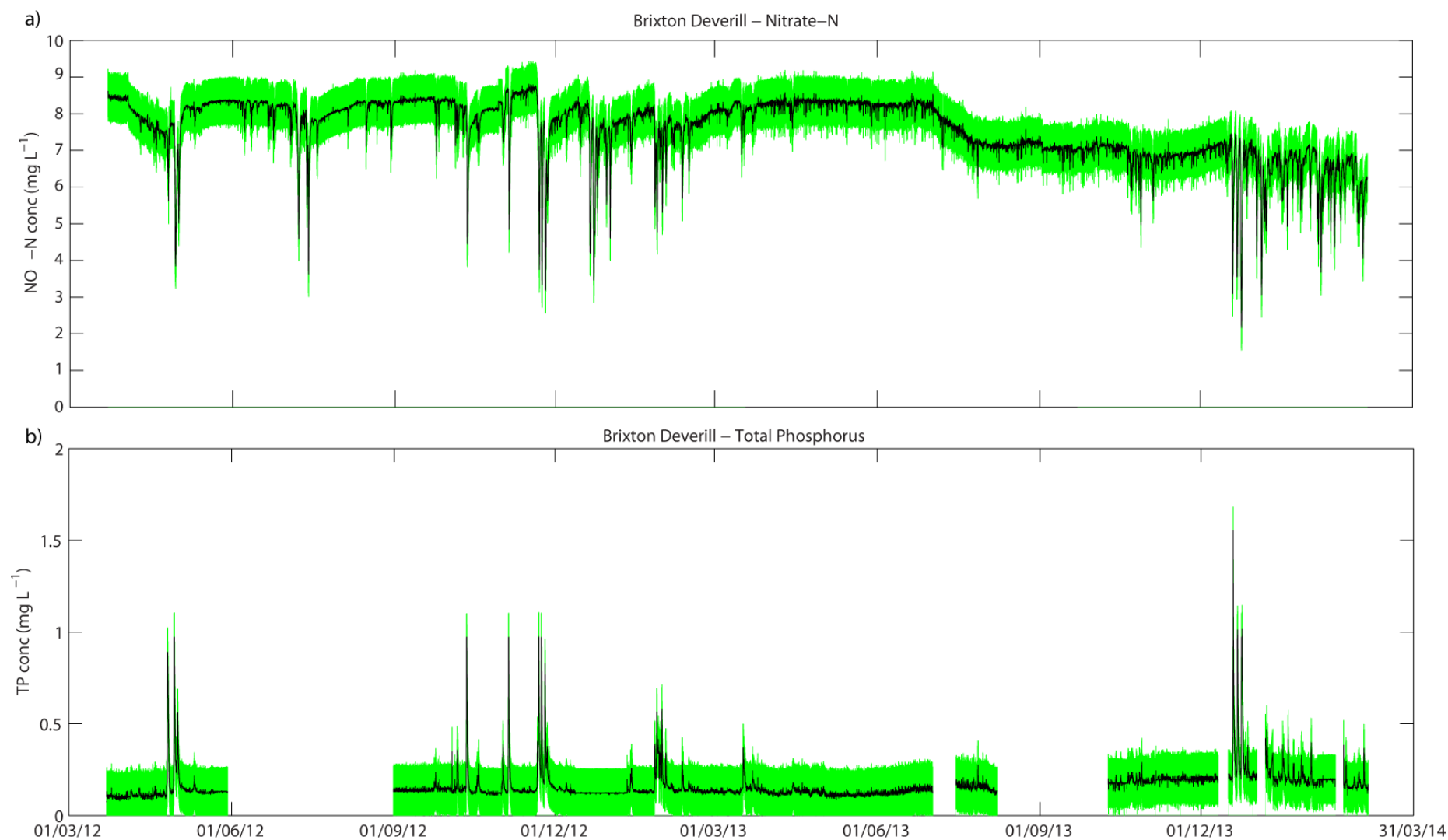
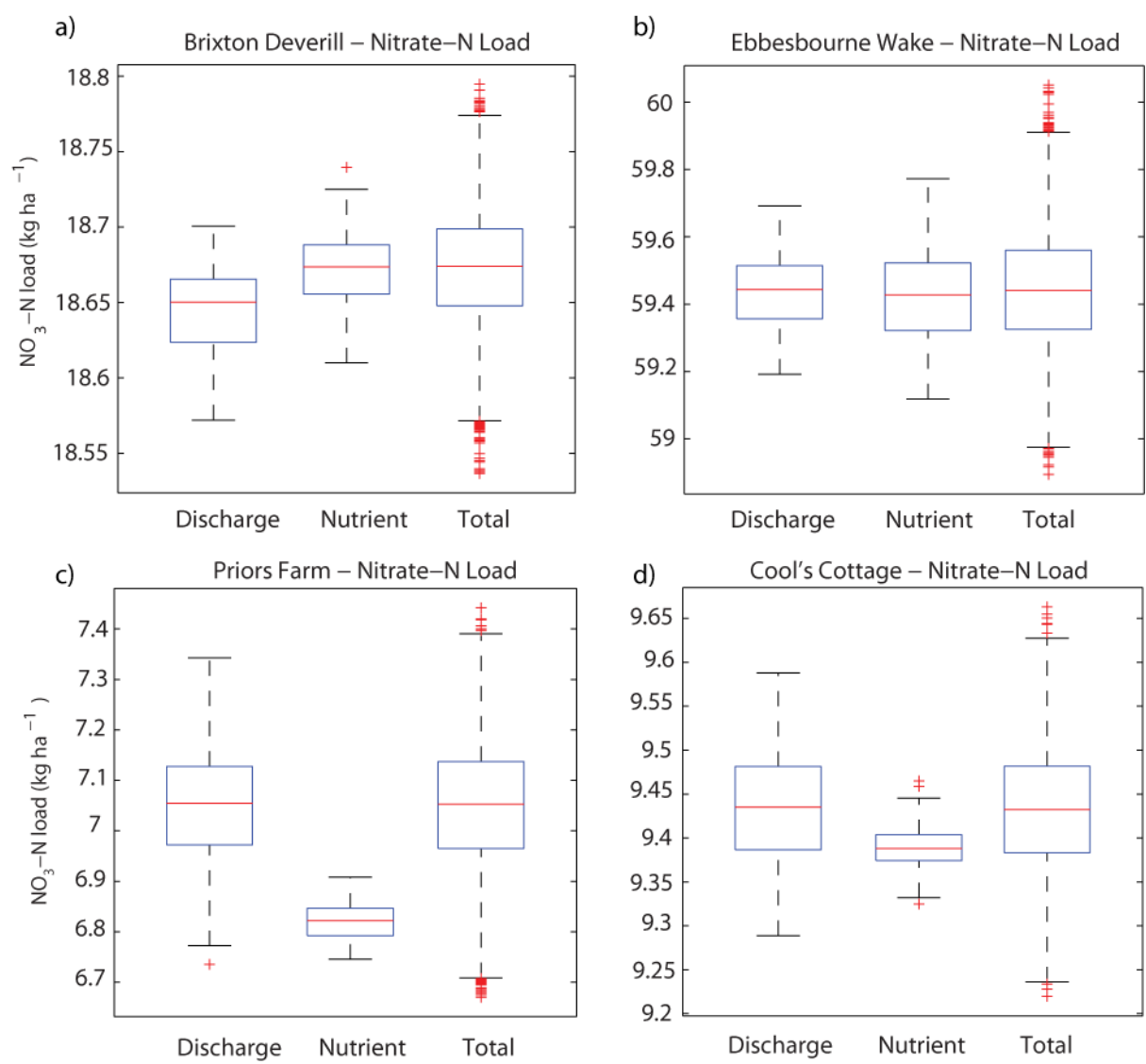


Figure 8: Plots showing total phosphorus concentrations from lab analysis of daily samples, where the green area shows the measurement uncertainty (10th-90th percentiles).



787

788 Figure 9: Plots showing a) NO₃-N and b) total phosphorus concentrations at Brixton Deverill from 30 min resolution in situ sensor data, where the green
789 area shows the measurement uncertainty (10th- 90th percentiles).



791

792

793

794

795

Figure 10: Boxplots showing the range of $\text{NO}_3\text{-N}$ loads including discharge uncertainty, nutrient uncertainty and total uncertainty for a) Brixton Deverill, b) Ebbesbourne Wake, c) Priors Farm and d) Cool's Cottage.

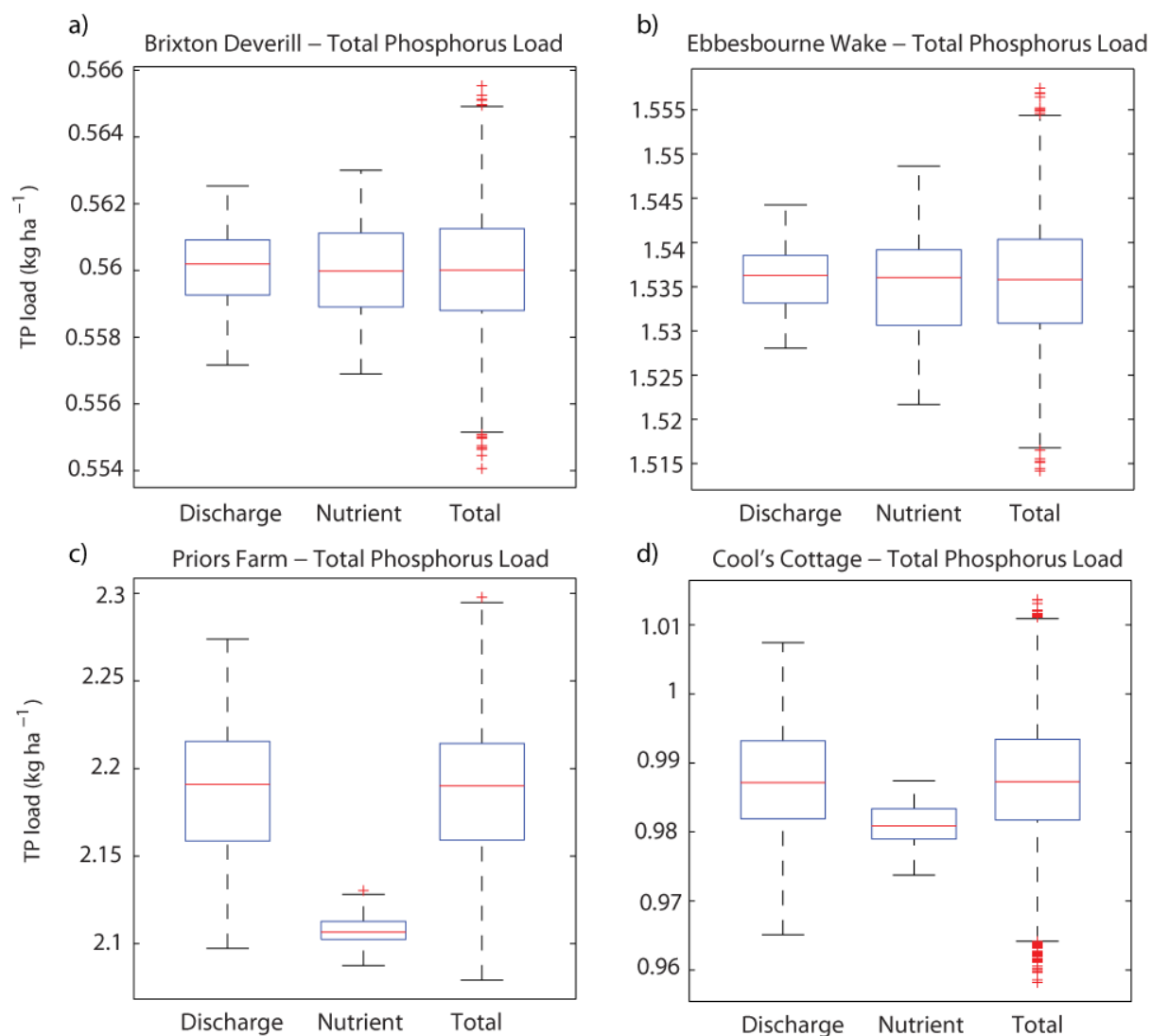


Figure 11: Boxplots showing the range of total phosphorus loads including discharge uncertainty, nutrient uncertainty and total uncertainty for a) Brixton Deverill, b) Ebbesbourne Wake, c) Priors Farm and d) Cool's Cottage.

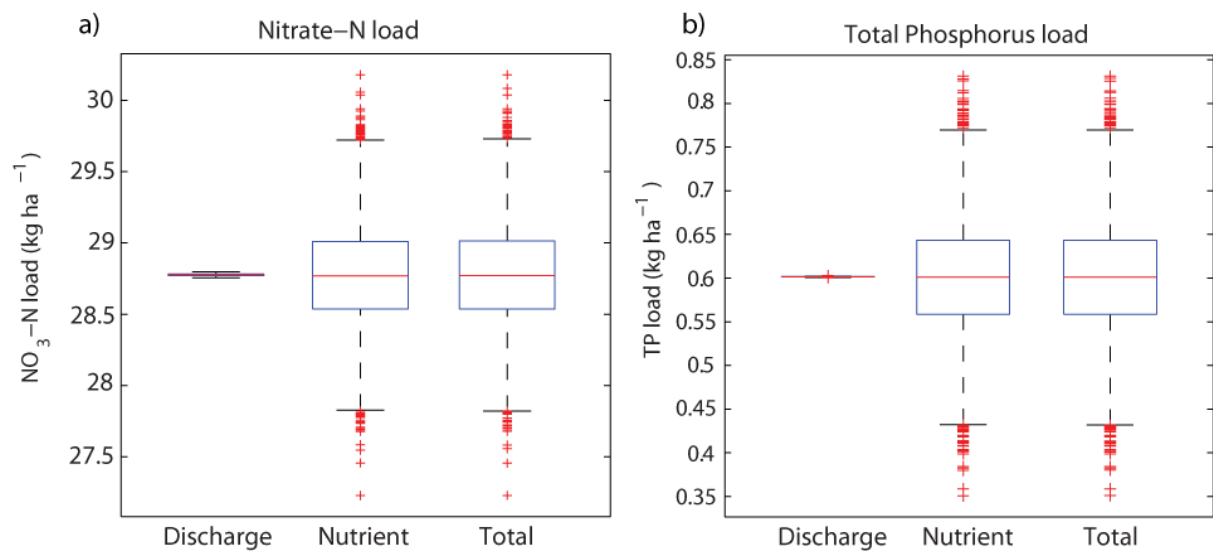


Figure 12: Boxplots showing the range of a) $\text{NO}_3\text{-N}$ and b) total phosphorus loads at Brixton Deverill using 30 min resolution sensor data, including discharge uncertainty, nutrient uncertainty and total uncertainty.